Clase 227 — GDPR y AI Act (EU)

Parte: 7 — Ética, Fairness y Privacidad · Fuente: Reglamento UE 2016/679 (GDPR) + Reglamento UE 2024/1689 (AI Act). Duración estimada: 75 min.

Clase 227 — GDPR y AI Act (EU)

> *Parte: 7 — Ética, Fairness y Privacidad · Fuente: Reglamento UE 2016/679 (GDPR) + Reglamento UE 2024/1689 (AI Act). Duración estimada: 75 min.*

Objetivo

Entender qué exige la regulación europea a un sistema de ML que toca datos personales o decisiones automatizadas. GDPR (en vigor desde 25-may-2018) regula el dato: bases legales, derechos del titular, DPIA. AI Act (Reglamento UE 2024/1689, escalonado 2024-2027) regula el sistema de IA por nivel de riesgo: prohibido / alto / limitado / mínimo. Aterrizamos ambas normas en un mini-toolkit programático que un equipo de datos puede ejecutar antes de poner un modelo en producción.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Identificar la base legal del Art. 6 GDPR aplicable a un tratamiento (consentimiento, contrato, interés legítimo, etc.) y distinguir las categorías especiales del Art. 9 (salud, biometría, raza).
- Implementar los derechos del titular más comunes: acceso, rectificación, supresión (Art. 17 — derecho al olvido) y portabilidad.
- Reconocer cuándo el Art. 22 GDPR (decisiones automatizadas con efectos significativos) exige supervisión humana y combinarlo con el Art. 14 del AI Act.
- Clasificar un caso de uso de IA en el nivel de riesgo del AI Act (prohibido, alto — Anexo III, limitado, mínimo) y listar las obligaciones que aplican.
- Generar una model card mínima y un checklist DPIA programático como artefactos de compliance.

Temas

#	Tema	Por qué importa
1	Bases legales (Art. 6) y categorías especi	Sin base legal válida, el tratamiento es i
2	Derechos del titular (acceso, supresión, p	El "derecho al olvido" obliga a poder borr
3	DPIA (Art. 35) — Data Protection Impact As	Obligatoria para alto riesgo (perfilado ma
4	AI Act: pirámide de riesgo	Prohibido → alto → limitado → mínimo. Defi
5	Sistemas de alto riesgo (Anexo III)	CV-screening, crédito, biometría, educació
6	GPAI y modelos fundacionales	Transparencia + resumen de datos de entren

Definiciones y características

- Dato personal (GDPR Art. 4.1): toda información sobre una persona física identificada o identificable — incluye IP, cookie ID, hashes débiles.
- Categorías especiales (Art. 9): origen racial, opiniones políticas, religión, datos genéticos, biométricos para identificar, salud, orientación sexual. Requieren base reforzada (consentimiento explícito, interés público en salud, etc.).
- Base legal Art. 6: una de seis: (a) consentimiento, (b) contrato, (c) obligación legal, (d) interés vital, (e)

interés público, (f) interés legítimo.
- DPIA (Art. 35): análisis previo de riesgos para los derechos del titular; obligatorio en perfilado sistemático a gran escala, categorías especiales, vigilancia pública.
- Decisión automatizada significativa (Art. 22): el titular tiene derecho a no ser sujeto a una decisión basada solo en tratamiento automatizado con efectos jurídicos o significativos — se levanta con consentimiento explícito, ejecución contractual o ley, siempre con derecho a intervención humana.
- Sistema de IA de alto riesgo (AI Act Anexo III): biometría, infraestructura crítica, educación, empleo, servicios esenciales (crédito, seguros), aplicación de la ley, migración, justicia. Obligaciones: gestión de riesgos, gobernanza de datos, documentación técnica, registros (logs), transparencia al usuario, supervisión humana (Art. 14), robustez, conformidad CE.
- GPAI (General Purpose AI): modelos fundacionales reutilizables. Transparencia + resumen público de datos de entrenamiento; si superan 10^{25} FLOPs entran en "systemic risk" con red-teaming y reporte de incidentes.
- Multas: GDPR hasta 20 M€ o 4% revenue global; AI Act hasta 35 M€ o 7% (prácticas prohibidas).

Dataset / recursos

- Dataset sintético de decisiones de crédito (1000 filas) con campos personales (email, dni, age, salary, score, is_minority) generado en el notebook con np.random.default_rng(42) — sin datos reales.
- Librerías: numpy, pandas, scikit-learn, re (regex para PII).

Ejercicios

1. Clasificador de riesgo AI Act: is_high_risk_use_case("CV screening") → "alto"; "juego móvil de match-3" → "mínimo". Cubrir los 4 niveles con lookup table basada en Anexo III.
2. DPIA checklist: dado {"sensitive_categories": True, "automated_decisions": True, "scale": "large"}, listar las obligaciones GDPR aplicables (DPIA, DPO, consentimiento reforzado, etc.).
3. Right to be forgotten: implementar right_to_be_forgotten(df, user_id) que elimine al usuario y devuelva un registro de auditoría con timestamp + columnas afectadas.
4. Data minimization audit: detectar columnas que parecen email (r"[\w\.-]+@[\w\.-]+") o DNI (r"\d{8}[A-Z]"); sugerir hash o remoción.
5. Compliance report: pipeline de scoring de crédito que ejecuta los 7 chequeos del notebook (clasificación de riesgo, DPIA, model card, supervisión humana, minimización, auditoría de PII, registro de borrado) e imprime un reporte único.

Homework verificable

Notebook con:

1. Tomar un caso de uso real propio o público (ej.: recomendador de empleo, scoring de seguros) y clasificarlo en el AI Act con la función del ejercicio 1.
2. Producir una model card completa (intended use, training data summary, performance global y por grupo sensible, limitaciones, fecha, owner).
3. Ejecutar el DPIA checklist y listar las obligaciones GDPR aplicables al caso.
4. Implementar el right to be forgotten sobre un dataset de >10K filas y verificar que tras la deleción el usuario no aparece en ninguna columna (incluido model.predict).
5. Generar un compliance report Markdown con todas las secciones (riesgo AI Act, DPIA, model card, PII audit, human-in-the-loop).

Criterio de aceptación: el reporte ejecuta sin errores, identifica correctamente el nivel de riesgo del caso, lista al menos 5 obligaciones GDPR justificadas, y demuestra el borrado de un usuario con auditoría.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
"Tengo consentimiento, ya puedo hacer todo	Consentimiento debe ser libre, específico,
Borrar al usuario de la tabla users pero s	El derecho al olvido aplica a todo el ecos
"Es solo un prototipo, GDPR no aplica"	GDPR aplica desde el primer dato personal
Clasificar un CV-screener como "riesgo lim	Anexo III lo lista explícitamente como alt
"Anonimicé los datos" pero quedan IPs y us	Pseudoanonimización ≠ anonimización. Si se
Modelo en producción sin human override pa	Art. 22 GDPR + Art. 14 AI Act exigen super

Preguntas frecuentes

Si mi empresa está fuera de la UE, ¿GDPR/AI Act aplican?

Sí, si tratás datos de personas en la UE (GDPR Art. 3, alcance extraterritorial) o si ofrecés un sistema de IA cuyo output se usa en la UE (AI Act Art. 2). Es el mismo principio del CCPA / LGPD.

¿GDPR me obliga a explicar mi modelo?

Hay debate. Wachter, Mittelstadt & Floridi (2017) argumentan que el Art. 22 + Recitales 71 garantizan información significativa sobre la lógica, no una explicación post-hoc tipo SHAP. En la práctica: documentación de la lógica + derecho a intervención humana + posibilidad de contestar la decisión.

¿Qué hago si entreno un LLM con datos scrapeados de internet?

Riesgo legal real (caso Clearview AI multado por la AEPD y otras DPAs). El AI Act prohíbe el scraping indiscriminado de imágenes faciales (Art. 5). Para texto, depende de base legal e intereses; minimizá, documentá fuentes, ofrecé opt-out.

¿Cuándo debo registrar un DPO (Data Protection Officer)?

Obligatorio si: autoridad pública, tratamiento a gran escala de categorías especiales, o monitoreo sistemático a gran escala (Art. 37). En la práctica, casi cualquier producto de ML con datos personales a escala lo requiere.

¿Cuándo entran en vigor las obligaciones del AI Act?

Escalonado: prohibiciones desde feb-2025, GPAI desde ago-2025, alto riesgo desde ago-2026/2027 según anexo. Sanciones empiezan a aplicarse con cada hito.

Referencias

- Reglamento UE 2016/679 (GDPR) — texto consolidado en EUR-Lex.
- Reglamento UE 2024/1689 (AI Act) — texto consolidado en EUR-Lex.
- European Data Protection Board — guidelines y decisiones.
- Council of Europe Framework Convention on AI (2024) — primer tratado internacional vinculante sobre IA.

- Wachter, S., Mittelstadt, B., Floridi, L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation (IDPL, 2017). <https://academic.oup.com/idpl/article/7/2/76/3860948>.
- Mitchell, M. et al. Model Cards for Model Reporting (FAT* 2019) — base de la model card que implementamos.

Siguiente clase

Clase 228 — Reproducibilidad: seeds, lock files, versionado de datasets

Apéndice: notebook (primer bloque)

Esta clase es legal/operativa. El notebook implementa un compliance toolkit que un equipo de datos puede correr antes de pasar un modelo a producción. Solo numpy, pandas, scikit-learn, re. Seed 42. Sin datos reales — generamos un sintético de decisiones de crédito.

```
import re
import json
import numpy as np
import pandas as pd
from datetime import datetime, timezone
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

rng = np.random.default_rng(42)
n = 1000

df = pd.DataFrame({
    'user_id':   np.arange(n),
    'email':     [f'user{i}@example.com' for i in range(n)],
    'dni':       [f'{rng.integers(10_000_000, 99_999_999)}{chr(rng.integers(65, 91))}' for _ in range(n)],
    'age':       rng.integers(18, 80, n),
    'salary':    rng.normal(35_000, 12_000, n).clip(8_000, 200_000).round(),
    'is_minority': rng.integers(0, 2, n),
    'past_default': rng.integers(0, 2, n),
})
df['approved'] = ((df['salary'] > 28_000) & (df['past_default'] == 0)).astype(int)
df['approved'] = np.where(rng.random(n) < 0.1, 1 - df['approved'], df['approved'])
df.head(3)
```

Archivos complementarios

- notebook.ipynb